

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-249331

(43)Date of publication of application : 27.09.1996

(51)Int.Cl.

G06F 17/27

(21)Application number : 07-049805

(71)Applicant : SHARP CORP

(22)Date of filing : 09.03.1995

(72)Inventor : TOKUNAGA SHINJI

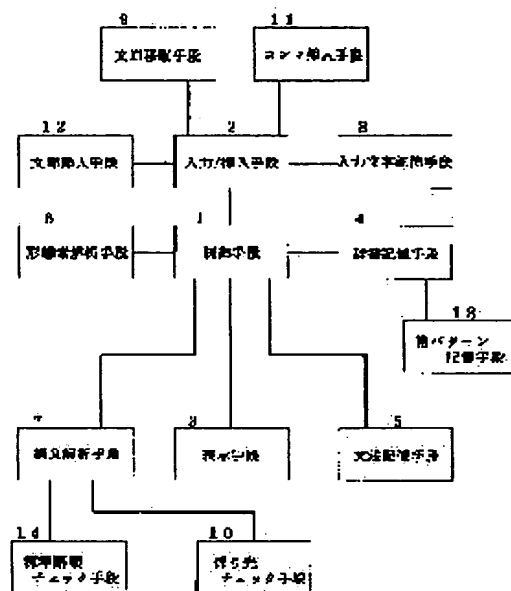
(54) DOCUMENT PROCESSOR

(57)Abstract:

PURPOSE: To accurately transmit the intention of a writer to a reader by judging whether or not there is a paragraph whose modification destination is not determined uniquely and moving a vague paragraph to a position where the vagueness of its modification destination is solved when the vague paragraph is present.

CONSTITUTION: A morpheme analyzing means 6 sections sentences into paragraphs by using a dictionary in a dictionary storage means 4 and grammar in a grammar storage means 5.

A syntax analyzing means 7 generates a syntax analytic tree by using the dictionary in the dictionary storage means 4, case patterns in a case pattern storage means 13, and the grammar in the grammar storage means 5. A modification destination check means 10 detects a sentence whose modification destination is vague by using the syntax analytic tree and a display means 3 displays it. Further, the sentence is moved by a paragraph moving means 9 to a position where the vagueness is eliminated or reduced and punctuation marks or commas are inserted by a comma inserting means 11. Further, a standard word order check means 14 checks whether or not the sentences appear in the order of case element standards and a paragraph moving means 9 moves paragraphs.



LEGAL STATUS

[Date of request for examination] 15.12.1998

[Date of sending the examiner's decision of rejection] 09.10.2001

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3420369

[Date of registration] 18.04.2003

[Number of appeal against examiner's decision of rejection] 2001-19887

[Date of requesting appeal against examiner's] 07.11.2001

decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-249331

(43) 公開日 平成8年(1996)9月27日

| | | | | |
|---------------------------|------|---------|---------------|---------|
| (51) Int.Cl. ⁵ | 識別記号 | 庁内整理番号 | F I | 技術表示箇所 |
| G 0 6 F 17/27 | | 9288-5L | G 0 6 F 15/20 | 5 5 0 E |
| | | 9288-5L | | 5 5 0 J |

審査請求 未請求 請求項の数4 O L (全 8 頁)

(21) 出願番号 特願平7-49805

(22) 出願日 平成7年(1995)3月9日

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 徳永 信治

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

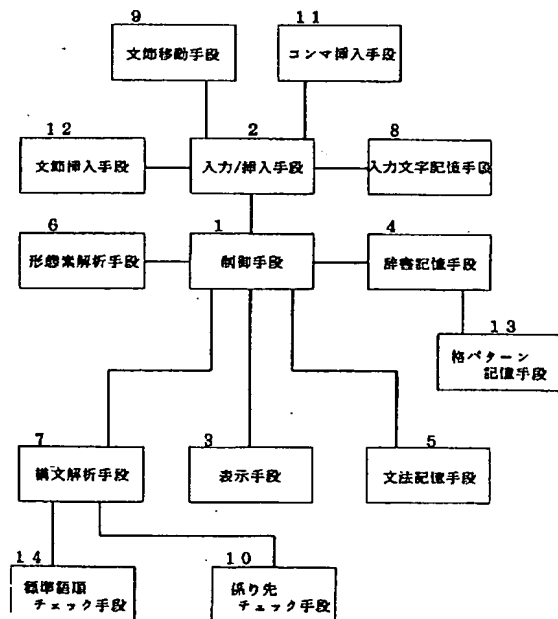
(74) 代理人 弁理士 梅田 勝

(54) 【発明の名称】 文書処理装置

(57) 【要約】

【目的】 構文解析を行い、係り先が一意に決まらない文節を指摘し、適切な位置への文節の移動、またはコンマ挿入を行う。また、格パターン辞書を用いて、標準の格要素の出現順でない文節の指摘を行い、文節の移動を行うことが可能である。

【構成】 形態素解析手段6で形態素解析、構文解析手段7で構文解析を行い、係り先チェック手段10で係り先が曖昧であると判断した場合に、文節移動手段9またはコンマ挿入手段11によって曖昧性を減少、解消する。



【特許請求の範囲】

【請求項 1】日本語を入力／編集する編集手段と、前記編集手段によって入力された日本語文を記憶する記憶手段と、前記日本語文を表示する表示手段と、辞書や文法を記憶する辞書記憶手段と、前記辞書記憶手段に記憶された辞書や文法に基づいて形態素解析を行う形態素解析手段、構文解析を行う構文解析手段を備えた文書処理装置において、

前記構文解析手段の結果から係り先が一意に決まらない文節の有無を判断する係り先チェック手段と、前記係り先チェック手段で曖昧な文節が存在する場合に、係り先の曖昧性を解消できる位置に該文節を移動する文節移動手段を備えることを特徴とする文書処理装置。

【請求項 2】日本語を入力／編集する編集手段と、前記編集手段によって入力された日本語文を記憶する記憶手段と、前記日本語文を表示する表示手段と、辞書や文法を記憶する辞書記憶手段と、前記辞書記憶手段に記憶された辞書や文法に基づいて形態素解析を行う形態素解析手段、構文解析を行う構文解析手段を備えた文書処理装置において、

前記構文解析手段の結果から係り先が一意に決まらない文節の有無を判断する係り先チェック手段と、前記係り先チェック手段で曖昧な文節が存在する場合に、係り先の曖昧性を解消できる位置に読点またはコンマを挿入するコンマ挿入手段を備えることを特徴とする文書処理装置。

【請求項 3】前記辞書記憶手段に、文節の共起情報、用言の格パターンを格納し、前記係り先チェック手段において、前記辞書記憶手段に格納された文節の共起情報を参照することによって、係り先のチェックを行うことを特徴とする請求項 1 または 2 に記載の文書処理装置。

【請求項 4】前記辞書記憶手段に、用言の格パターンを格納し、前記係り先チェック手段において、解析結果と前記辞書記憶手段に格納された格パターンを照合し、前記辞書記憶手段に格納された格パターンの順序に、入力文の名詞句文節を移動させる文節移動手段を備えることを特徴とする文書処理装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、日本語の校正／推敲を行う日本語ワードプロセッサ、機械翻訳装置などに関する。

【0002】

【従来の技術】従来、日本語校正／推敲を行う文書処理装置においては、日英機械翻訳装置などでは必ず行われる構文解析処理を行うものは少なく、ほとんどが形態素解析処理のみで、誤りを指摘するものがほとんどであった。

【0003】例えば、特開平 4-213155 に開示されている発明においては、形態素解析処理を行い、より確からし

い単語の並びが存在する場合に限って誤字候補を出力するというものが記されている。単語の品詞その他の情報から表層の並びの出現確率を定義したコストテーブルを利用して、単語辞書に登録がなかった未登録語および接続コストが閾値よりも低い同音語がある場合、つまり文法的により確からしい単語の並びが存在する場合に限って誤字候補を出力するようにしている。

【0004】また、特開昭 62-150461、特開昭 62-150462 には呼応関係を有する副詞と助動詞の組を記憶しておき、それに該当する組み合わせの表現がない場合に、検出／修正を行うものである。これによって、呼応関係にある語の用法に誤りがある可能性がある部分を検出／修正することが可能となっている。

【0005】特開昭 62-209659 においては、形態素解析、構文解析を行い、制約条件の付いた文法を用いて、その制約条件に該当する箇所を指摘することによって、校正を支援している。

【0006】

【発明が解決しようとする課題】しかしながら、形態素解析のみを行って校正／修正を行う場合においては、係り受け関係は無視されているため、誤字の訂正は可能ではあるが、係り受け関係が明確にならず、校正／修正ができないことがある。

【0007】例えば次のような 2 つの文を例にとる。

【0008】(1)「彼は休憩しているところを見た。」

(2)「彼が休憩しているところを見た。」

この 2 つの文では「彼」「休憩している」のか「見た」のかは曖昧にとれる。一般的に (1) の文では「彼は (誰かが) 休憩しているところを見た。」というように「彼」が遠くの「見た」にかかるのが自然であり、一方 (2) の文では、

「(誰かが) 彼が休憩しているところを見た。」

というように「彼」が近くの「休憩している」に係ると理解するのが自然である。

【0009】そこで本発明では、構文解析を行うことにより、係り受け関係を解析し、係り受け関係が曖昧である場合に、曖昧性を解消する位置に文節を移動したり、係り受けの係り文節の直後に読点やコンマを挿入することによって、係り受けの曖昧性を解消し、読みやすい日本語文にする。

【0010】例えば、文節移動による曖昧性の解消では、(1a)「休憩しているところを彼は見た。」となり、また、読点による曖昧性の解消では、(1b)「彼は、休憩しているところを見た。」となる。

【0011】この場合に、読点やコンマについては、必要以上にあるとかえって読みにくい文章になってしまうので、構文解析を行って、曖昧性がある場合のみを挿入対象とする。

【0012】また、構文解析を行う場合においても、従

来の技術のように文法に制約条件を入れるという特殊な処理を行うのではなく、1つの用言に係っているすべての要素を認識し、かつ辞書情報として持っている格パターンと比較する格解析を行い、語順のチェックなどを行う。

【0013】

【課題を解決するための手段】

〔請求項1の手段〕日本語を入力／編集する編集手段と、前記編集手段によって入力された日本語文を記憶する記憶手段と、前記日本語文を表示する表示手段と、辞書や文法を記憶する辞書記憶手段と、前記辞書記憶手段に記憶された辞書や文法に基づいて形態素解析を行う形態素解析手段、構文解析を行う構文解析手段を備えた文書処理装置において、前記構文解析手段の結果から係り先が一意に決まらない文節の有無を判断する係り先チェック手段と、前記係り先チェック手段で曖昧な文節が存在する場合に、係り先の曖昧性を解消できる位置に該文節を移動する文節移動手段を備える。

【0014】〔請求項2の手段〕日本語を入力／編集する編集手段と、前記編集手段によって入力された日本語文を記憶する記憶手段と、前記日本語文を表示する表示手段と、辞書や文法を記憶する辞書記憶手段と、前記辞書記憶手段に記憶された辞書や文法に基づいて形態素解析を行う形態素解析手段、構文解析を行う構文解析手段を備えた文書処理装置において、前記構文解析手段の結果から係り先が一意に決まらない文節の有無を判断する係り先チェック手段と、前記係り先チェック手段で曖昧な文節が存在する場合に、係り先の曖昧性を解消できる位置に読点またはコンマを挿入するコンマ挿入手段を備える。

【0015】〔請求項3の手段〕前記辞書記憶手段に、文節の共起情報、用言の格パターンを格納し、前記係り先チェック手段において、前記辞書記憶手段に格納された文節の共起情報を参照することによって、係り先のチェックを行う。

【0016】〔請求項4の手段〕前記辞書記憶手段に、用言の格パターンを格納し、前記係り先チェック手段において、解析結果と前記辞書記憶手段に格納された格パターンを照合し、前記辞書記憶手段に格納された格パターンの順序に、入力文の名詞句文節を移動させる文節移動手段を備える。

【0017】

〔作用〕請求項1にかかる発明によれば、編集手段によって入力された入力文に対して、構文解析手段を行い、その結果を係り先チェック手段で判断し、曖昧な文節が存在する場合、その旨を表示し、係り先候補のうち、ユーザがより遠くの文節に係る意味でその文を書いている場合に、ユーザがわざわざ範囲指定や移動の操作を行うことなく、文節移動手段によって効率よくより遠くの受け文節の直前へ該文節を移動することができる。

【0018】請求項2にかかる発明によれば、上記の請求項1では文節を移動することによって、曖昧性を解消していたが、コンマ挿入手段によって、係り文節の直後へカーソル移動などの操作をせずに、読点やコンマを挿入することができる。

【0019】請求項3にかかる発明によれば、文節の共起情報、用言の格パターンを利用して、請求項1および請求項2の係り先チェック手段において、文節の共起情報を利用することによって、係り受けのチェックを行い、文節の移動または読点／コンマの挿入を行い、係り受けの曖昧性を解消する。

【0020】請求項4にかかる発明によれば、前記辞書記憶手段に、用言の格パターンを格納し、前記係り先チェック手段において、解析結果と前記辞書記憶手段に格納された格パターンを照合し、前記辞書記憶手段に格納された格パターンの順序に、入力文の名詞句文節を移動させる文節移動手段を備える。

【0021】

【実施例】以下、本発明による文書処理装置の一実施例を図面を用いて説明する。図1は、本発明における文書処理装置の構成の一例を示した図である。1は制御手段、2は入力／編集手段、3は表示手段、4は辞書記憶手段、5は文法記憶手段、6は形態素解析手段、7は構文解析手段、8は入力文字記憶手段、9は文節移動手段、10は係り先チェック手段、11はコンマ挿入手段、12は文節挿入手段、13は格パターン記憶手段、14は標準語順チェック手段を示している。

【0022】制御手段1の制御により、入力手段2から入力された入力文は、入力文字記憶手段8に格納される。これと同時に、表示手段3にもこの内容は表示される。形態素解析手段6では、辞書記憶手段4の辞書や文法記憶手段5の文法を用いて、文を文節ごとに区切る。

【0023】構文解析手段7では、辞書記憶手段4の辞書や格パターン記憶手段13の格パターンおよび文法記憶手段5の文法を用いて、構文解析木を作成する。構文解析木から、係り先チェック手段10により、係り先の曖昧な文を検出し、表示手段3において、表示を行う。また、曖昧性を解消または減少する位置へ文節移動手段9により移動したり、コンマ挿入手段11で、読点やコンマを挿入したりする。

【0024】さらに、構文解析木と格パターン記憶部13の格パターンから、標準語順チェック手段14により、格要素標準の順に出現しているかどうかをチェックし、文節移動手段9によって移動を行う。

【0025】図2は、本発明の処理を示すフローチャートである。図9は画面表示イメージの例を示す図である。まず、構成対象の文章を表示するメインウィンドウがある。全文が表示できない場合は、スクロールなどの機能により、全体を確認することが可能である。

【0026】本発明では、基本的に一文毎の処理である

ので、処理対象としている一文と、指摘内容を表示するウィンドウを用意する。図9では、「指摘対象文を示すウィンドウ」として示す。図9(a)、図9(b)において、下線を付した文節は、文節を移動する方がよい文節を示し、空枠は移動すべき位置を示している。

【0027】また、校正操作をメニューにしたウィンドウを「校正操作選択メニューウィンドウ」として示す。システムの指摘通りにするかどうかを選択する「●」

「○」はそれぞれ選択、非選択を示し、マウスなどで変更／選択できるようになっている。

【0028】さらに図9には図示していないが、問題となる理由(内容)を表示するウィンドウを表示させてもよい。例えば図9(a)であれば「係り先の曖昧な文節があります。曖昧さを軽減できる位置に該当の文節を移動しますか?」という表示を、また図9(b)であれば「用言に対する格助詞の順序が標準と異なります。標準の語順に変更しますか?」と表示させる。

【0029】図2はメインフローである。ステップS1では、入力文一文が入力文字記憶手段8に読み込まれる。それと同時に表示手段3に表示が行われる。ステップS2では形態素解析手段6で、辞書記憶手段4の辞書や文法記憶手段5の文法を用いて、文節を作成する。ステップS3では、形態素解析レベルで行える校正支援を行う。例えば、表記の間違い指摘や語尾の不統一の指摘などを行う。

【0030】ステップS4において、ステップS3で行われた校正支援での指摘に基づいて、入力文が変更されたかどうかを判断し、変更された場合は、再び形態素解析を行うために、ステップS2に戻る。変更が行われなかった場合にはステップS6に処理を進める。

【0031】ステップS6では、構文解析処理を行う。構文解析処理のアルゴリズムは特に特別なものではないが、格パターンを用いた格解析／意味解析を行う。

【0032】ステップS7において、ステップS6での結果を利用して校正支援を行う。詳細は図3、図4のフロー、図5の例文を用いて説明する。なお、この図3、図4で示す処理の順序は問題ではない。

【0033】次に図3を用いて、係り受けに曖昧性があった場合の処理を示す。ステップS11において、図5(1-3)のような文では、構文解析の結果は、図7に示すように2種類の係り受けの可能性がある。図7は依存構造であり、受けの語の下に係りの語を並べる構造である。図7(1-3a)は、「本を」と「昨日」が「読んだ」に、「買った」が「本を」に係ることを示し、(1-3b)は、「昨日」が「買った」に、「買った」が「本を」に、「本を」が「読んだ」に係ることを示している。この構造を同じように示したのが図6(1-3)である。係り先チェック手段10を用いて、この構造から2カ所にかかる可能性がある文節を探し、「昨日」を見つける。2カ所以上に係る可能性がある文節があった場合にはステップ

S12へ、ない場合はメインフローへ処理に戻る。

【0034】請求項3の実施例の場合、2つ以上ある係り先のうち、どちらに係る可能性が高いかを、格パターン記憶手段13の格パターンを用いて、格解析および意味解析を行い判断する。

【0035】図5(1-1a)では、図6(1-1a)に示すように「ギターを」は、より遠くの「弾いている」に、図5(1-1b)では、図6(1-1b)のように、より近くの文節「唄っている」に係る。同様に、図5(1-2a)および(1-2b)の例では、(1-2a)の例が、より遠くの文節に係っている。請求項3ではこのステップで、係り先の曖昧な文節があり、かつ係り先の可能性がある文節のうち、より遠い方に係っているとシステムが判断した場合、ステップS32へ、それ以外の場合はメインフローに戻る。

【0036】つまり、図5の(1-1a)および(1-2a)の文では、ステップS12へすみ、図5(1-1b)および(1-2b)の文はメインフローに戻る。また、(1-3)の例では「昨日」が「買った」に係るのか「読んだ」に係るのか判断できない。さらに図5(1-4)、図6の(1-4)でも「赤い」が「会社」に係るのか「車」に係るのか判断できない(「赤い会社」なるものがないという知識を格納していれば判断できる)。これらの場合も、ステップS12でユーザの判断にゆだねることになる。

【0037】ステップS12では、係り先の曖昧な文節の特徴表示(例えばカラー表示、アンダーライン表示など)と、移動先の位置を表示する。図5の(1-1a)の文の場合、図9の(1-1a)の「指摘対象文を示すウィンドウ」のような表示を行う。また、「校正操作メニューウィンドウ」のように、選択項目を表示する。ここで、文節移動を行うか、読点(またはコンマ)を挿入するか、そのままの選択肢を表示し、ステップS13へ進む。

【0038】ステップS13では、ステップS12で表示の選択肢からユーザが選択し、ステップS14に進む。

【0039】ステップS14では、ステップS13で文節移動を選択した場合には、特徴表示している文節を移動先の位置に移動する。移動先は、係り先の文節の直前である。また、読点挿入を選択した場合は、特徴表示した文節の直後へ読点(またはコンマ)を挿入する。それぞれの処理終了後メインフローへ戻る。

【0040】次に、格解析を用いて語順のチェックを行い、語順を変更する処理を説明する。ステップS21では、構文解析を行った結果、各用言にどんな文節がどんな順で係っているのか解析される。これを格パターンデータと比較することにより、標準的な語順であるかどうか判断される。例えば、図5(2-1)および(2-2)では、用言「書いた」には「本を」と「私が」が、この順で係っている。また、用言「見なす」には、「正解と」と「これを」とが、この順で係っている。この用言には、格パターンデータとして図8のようなものがある。格パ

ターンの各要素の順を標準語順とすると、両者とも異なっている。このように標準語順と異なっている場合は、ステップS22へ進み、異なっていない場合はメインフローに戻る。

【0041】ステップS22では、標準語順にするための文節移動の表示を行う。つまり、図9の(2-1)の「指摘対象文を示すウインドウ」のように、移動する文節と移動先を表示する。格要素が3つ以上ある場合は、ステップS21～ステップS24が繰り返される。

【0042】語順の変更を、格要素である文節の移動と考えて、標準の語順で用言に最も近い格要素からチェックし、入力文でその格要素と用言との間に他の格要素が入っていると、用言の直前に移動先を設定する。繰り返される場合は、その次に用言に近い各要素をチェックし、先に移動し終えた用言に最も近い格要素の直前に移動先を設定する。その後ステップS23へ進む。

【0043】ステップS23では、ユーザは移動するかどうかを選択する。移動を行う場合には、ステップS24に進み、そうでない場合はメインフローに戻る。

【0044】ステップS24では、特徴表示している文節を移動先の位置へ移動する。そしてメインフローに戻る。

【0045】ステップS7では、ステップS3の従来の校正支援機能による校正の結果およびステップS6の本発明による校正支援機能による校正の結果で、入力文字記憶手段8の内容および表示手段3の内容が変更される。

【0046】ステップS8で、次の文があればステップS1に戻り、なければ、処理を終了する。

【0047】

【発明の効果】

【請求項1の効果】本発明によれば、係り先が曖昧な文節に対して、より遠い方の係り先に係る方が、ユーザの意図するない用である場合に、該文節を係り先の直前に移動し、曖昧性を解消もしくは減少させた文に校正することが可能であり、読み手に書き手の意図を正確に伝えることが可能となる。

【0048】【請求項2の効果】また、文節の移動の代わりに、該文節の直後に読点かコンマを挿入することによって、読み手に該文節が遠方に係ることを示している

ことを暗に伝えることができ、曖昧性を軽減することが可能となる。

【0049】【請求項3の効果】さらに、格解析／共起情報を用いることによって、正確に係り受けを判断することが可能となる。

【0050】【請求項4の効果】日本語において、用言の格要素の語順は比較的自由であるが、特別な効果をねらう場合以外は標準的な語順がある程度決まっており、読み手に伝わりやすい。よって、標準的でない語順を指摘し、標準的な位置に移動することによって、読み手に伝わりやすい。

【図面の簡単な説明】

【図1】本発明の一実施例に係るブロック図である。

【図2】本発明に関するメインフロー図である。

【図3】本発明の請求項1または2の処理に関するフロー図である。

【図4】本発明の請求項4の処理に関するフロー図である。

【図5】入力文の例を示す図である。

【図6】入力文の例の係り受け構造を示す図である。

【図7】入力文の例の構文解析結果を示す図である。

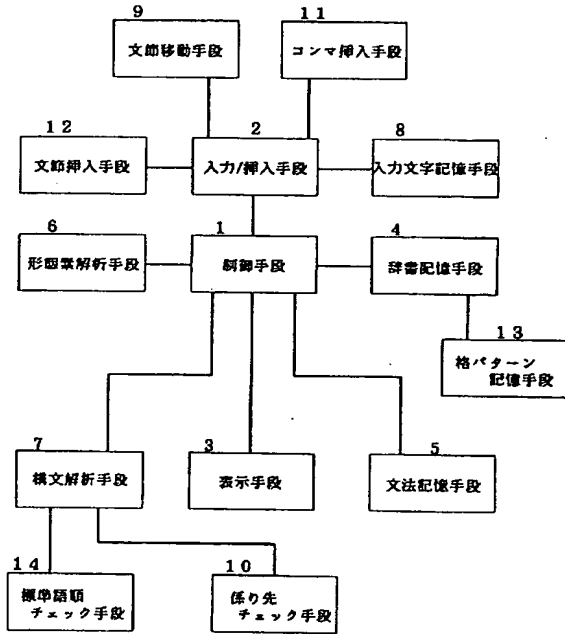
【図8】用言辞書の格パターンの例を示す図である。

【図9】本発明の一実施例の画面表示例を示す図である。

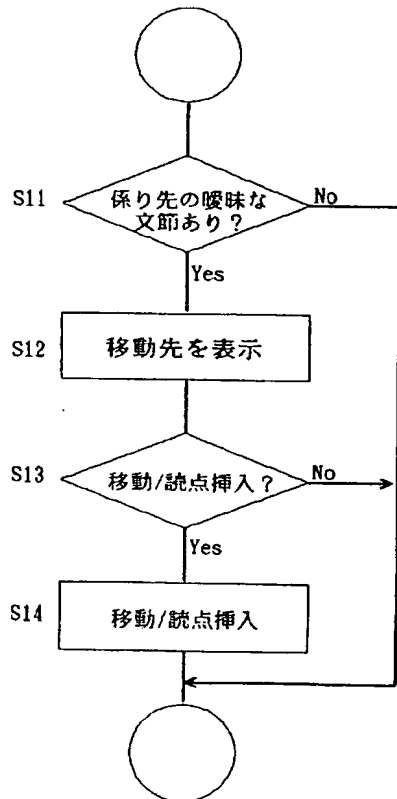
【符号の説明】

- 1 制御手段
- 2 入力／挿入手段
- 3 表示手段
- 4 辞書記憶手段
- 5 文法記憶手段
- 6 形態素解析手段
- 7 構文解析手段
- 8 入力文字記憶手段
- 9 文節移動手段
- 10 係り先チェック手段
- 11 コンマ挿入手段
- 12 文節挿入手段
- 13 格パターン記憶手段
- 14 標準語順チェック手段

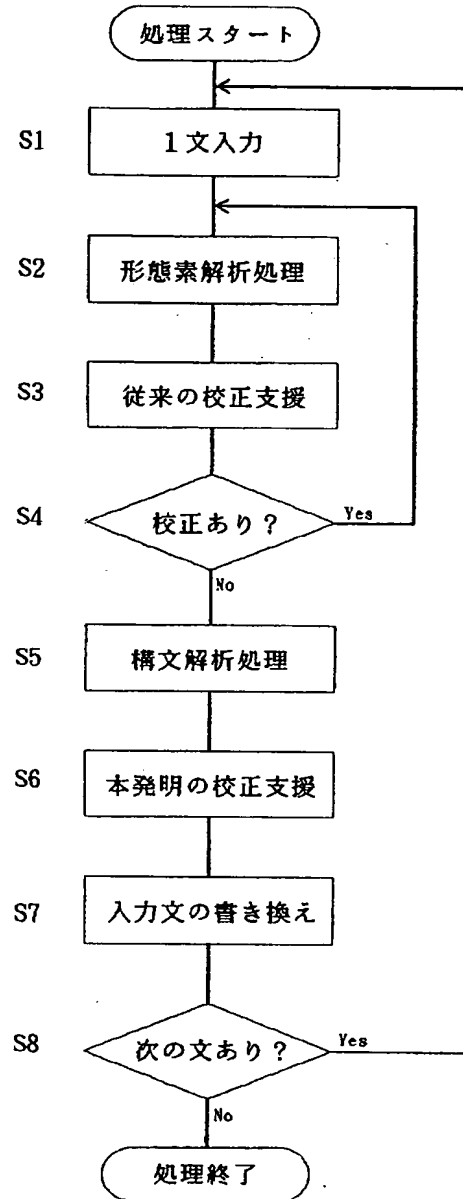
【図 1】



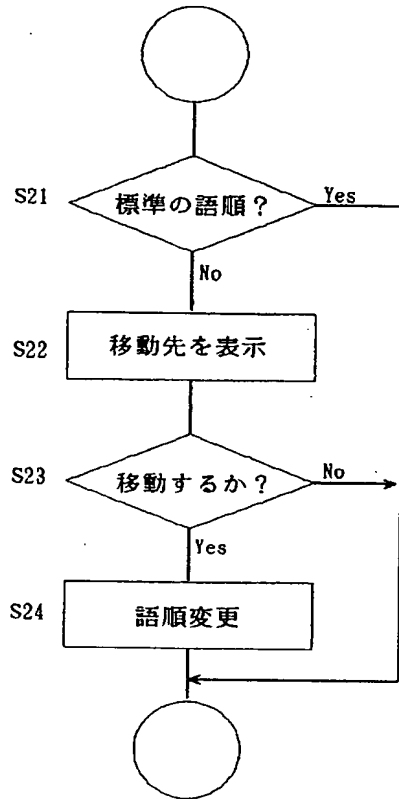
【図 3】



【図 2】



【図 4】



【図 5】

(1-1a) ギターを弾いている人が弾いている。

(1-1b) メロディーを弾いている人が弾いている。(1-3a)

(1-2a) 雨が降るまで降っていた。

(1-2b) 彼が帰るまで降っていた。

(1-3) 昨日買った本を読んだ。

(1-4) 赤い会社の車

(2-1) 本を私が書いた。

(2-2) 正解とこれをみなす。

昨日買った本を読んだ。

読んだ
└─ 本を
 └─ 買った
 └─ 昨日

(1-3b) 読んだ
└─ 本を
 └─ 買った
 └─ 昨日

【図 6】

(1-1a) ギターを 弾いている 人が 弾いている。

(1-1a') 弾いている 人が ギターを 弾いている。

(1-1b) メロディーを 弾いている 人が 弾いている。聞く

(1-2a) 雨が 降るまで 降っていた。

(1-2a') 降るまで 雨が 降っていた。

(1-2b) 彼が 降るまで 降っていた。

(1-3) 昨日 買った 本を 読んだ。

昨日 買った 本を 読んだ。
└─ 昨日 └─ 買った └─ 本を └─ 読んだ

(1-4) 赤い 会社の 車

赤い 会社の 車
└─ 赤い └─ 会社の └─ 車

読む 《人間》 が
《本》 を
来た 《人間》 が
《場所》 に、へ
会う 《人間》 が
《人間》 に、と

《人間》 が
《人間》 に
《情報》 を

書く 《人間》 が
《本》 を

見なす 《人間》 が
() を
() と

【図 9】

(全文表示ウィンドウ)

ギターを唄っている人が弾いている。ピアノは女性が弾いている。

(指摘対象文を示すウィンドウ)

ギターを 唄っている 人が 弾いている。

(校正操作選択ウィンドウ)

- 移動する
- 読点(コマ)挿入
- なにもしない

(指摘対象文を示すウィンドウ)

本を 私が 書いた。

(校正操作選択ウィンドウ)

- 移動する
- なにもしない